Cloning and sequencing of full-length cDNA encoding the precursor of human complement component C1r

Agnès JOURNET*‡ and Mario TOSI†

*D.R.F. Groupe Immunochimie (Unité INSERM U238), Centre d'Etudes Nucléaires de Grenoble, 85X, 38041 Grenoble Cedex, France, and †Unité d'Immunogénétique (Unité INSERM U276), Institut Pasteur, 75724 Paris, Cedex 15, France

The sequencing of human liver cDNA clones encoding the entire C1r precursor protein has confirmed the previously determined peptide sequence and has shown that there is a leader peptide which is 17 amino acids long. A residue tentatively identified as β-hydroxyaspartic acid [Arlaud, Willis & Gagnon (1986) Biochem. J., in the press] located in the C1̄r A-chain, within an epidermal-growth-factor consensus sequence, was found to be encoded as asparagine. Two sequence elements, tandemly located in the A-chain, are related to a sequence widespread among proteins which interact with C3b or C4b. Structural comparisons between different clones indicate that multiple polyadenylation sites are responsible for the length heterogeneity observed for C1r mRNA from liver and Hep G2 cells.

INTRODUCTION

The first enzymic event in the cascade of the classical pathway of complement is the activation of its first component, C1. C1 is a multimolecular complex consisting of two loosely interacting entities: C1q, the recognition subunit, and $C1r_2$–$C1s_2$, the catalytic one. $C1r_2$–$C1s_2$ is a tetrameric Ca^{2+}-dependent complex of C1r and C1s, two serine-proteinase zymogens. These zymogens are activated in sequence during the C1 activation (Reid & Porter, 1981; Cooper, 1985): first, the dimeric form of C1r autoactivates by proteolysis of a single Arg–Ile bond in each monomer (Arlaud & Gagnon, 1985a); $C1̄r_2$ then cleaves an Arg–Ile bond (Spycher et al., 1986) in each molecule of C1s, leading to C1̄s, which is able to proteolytically attack its substrates C4 and C2.

C1r plays a key role in these series of events, due to its autoactivation potential (Arlaud et al., 1980). It acts by transforming the activation signal (the fixation of C1q on activating reagents, e.g. immune complexes) into an enzymic activity. Up to now, most biochemical studies on human C1r have been performed on the protein purified from plasma pools. Its complete sequence has been determined since the present studies were initiated (Arlaud & Gagnon, 1985b; Arlaud et al., 1986b). The mature plasma form has an approximate M_r of 90000 and contains 688 amino acids. An M_r of about 76000 for the unglycosylated precursor has been estimated by cell-free translation of human liver mRNA, immunoprecipitation and SDS/polyacrylamide-gel-electrophoretic analysis (Tosi et al., 1985). Intracellular and secreted C1r have been studied by using cultured Hep G2 hepatoma cells (Reboul et al., 1986). These cells synthesize low amounts of C1r. Nevertheless, it has been possible to observe that the protein secreted by Hep G2 cells does not autoactivate, although it is synthesized in its dimeric form and is activated by exogeneous C1̄r. The isolation of cDNA and/or genomic clones

encoding human C1r should provide the opportunity to produce this protein under controlled conditions and to monitor the evolution of C1r properties along the steps of its synthesis and secretion.

As an initial step, we have characterized cDNA clones encoding human C1r and have deduced the entire sequence of the C1r precursor protein.

MATERIALS AND METHODS

Liver mRNA preparations, cell-free translation and immunoprecipitation

RNA extraction from post-mortem human liver, purification of polyadenylated RNA and its size fractionation on sucrose gradients have been already described (Tosi et al., 1986). The C1r mRNA enrichment was determined by cell-free translation and immunoprecipitation with rabbit antibodies to C1r (Tosi et al., 1986).

Construction of a C1r-enriched human liver cDNA library

Double-stranded cDNA was synthesized (Gubler & Hoffman, 1983) from sucrose-gradient-fractionated mRNA. cDNA was poly(dC)-tailed and annealed with poly(dG)-tailed pUC9 vector (Pharmacia/PL-Biochemicals; Vieira & Messing, 1982) as described by Maniatis et al. (1982). Competent *Escherichia coli* 5K bacteria (Hubacek & Glover, 1970) were prepared essentially as described by Hanahan (1983). An initial cDNA library (50000 independent transformants) was amplified approx. 500-fold. About 40000 ampicillin-resistant bacteria were screened with the synthetic oligodeoxynucleotide mixture shown in the Results section. This probe was synthesized and used in colony-filter hybridizations according to protocols that have already been described (Tosi et al., 1986).
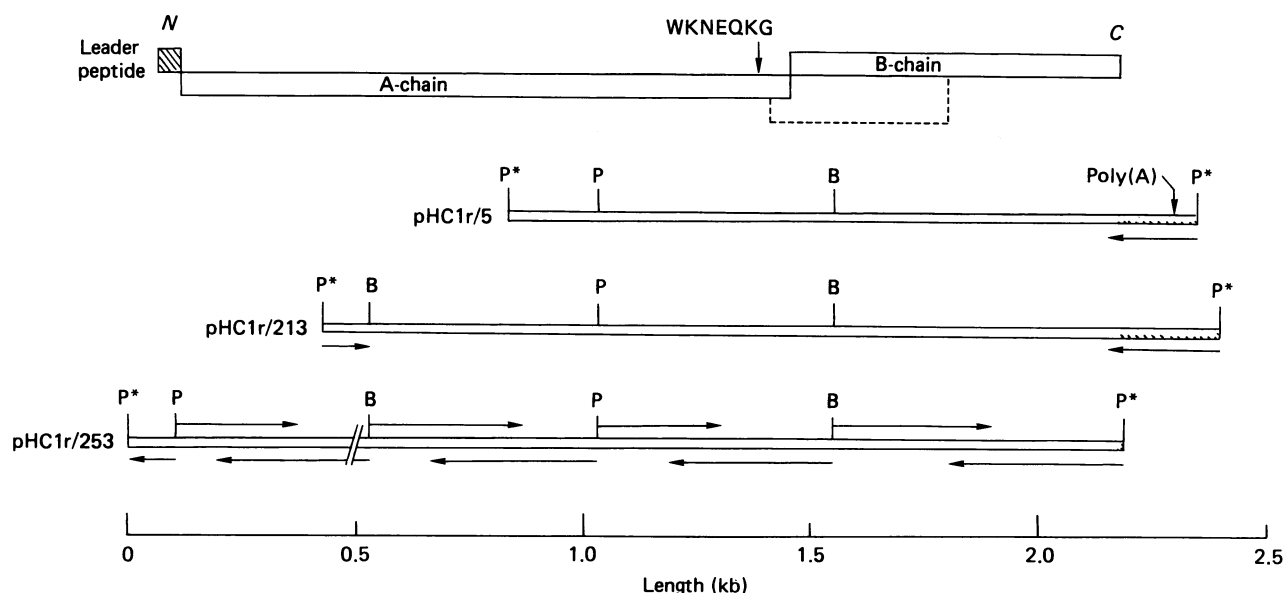
Fig. 1. Alignment of cDNA inserts with the human C1r precursor

The amino acid sequence used to design the synthetic oligonucleotide mixture is shown above the scheme of the C1r protein, and the disulphide linkage (Gagnon & Arlaud, 1985; Arlaud & Gagnon, 1985a) between the A- and the B-chains is indicated by a broken line. The hatched portion of the cDNA inserts represents 3' non-coding sequences. The restriction sites PstI (P) and BamHI (B) used for subcloning are shown together with the direction and extent of sequence information obtained from each fragment. Additional restriction fragments that have been sequenced are not indicated for sake of clarity. In particular, a subcloned 362-bp-long HindIII/SmaI fragment derived from the 5' extremity of the pHC1r/253 insert was used to confirm the nucleotide sequence of the leader peptide and of the portion of the 5' non-coding region shown in Fig. 2. Note that each insert is delimited by PstI sites produced by the cloning procedure (P*). The bar representing the pHC1r/253 clone is broken to indicate a 5 bp deletion.

Search for full-length clones

This original library was then enriched for long inserts by agarose-gel electrophoresis of the pool of recombinant plasmids which had been separately linearized with restriction endonucleases HindIII, SalI and EcoRI. Linearized plasmids with an insert longer than 1.5 kb were recovered on an NA 45 membrane (Schleicher and Schuell), eluted according to the manufacturer's instructions, pooled and re-ligated. Appropriate aliquots of these size-selected plasmids were used to transform competent bacteria as indicated above. A 200-bp-long 5' fragment, obtained from an initial clone pHC1r/5 (see Fig. 1), was radioactively labelled by nick-translation (Maniatis et al., 1982). Filters carrying a total of about 12000 colonies were prehybridized for 3–6 h at 42 °C, in 50% (v/v) formamide/5 × SSC (75 mM-sodium citrate/750 mM-NaCl)/5 × Denhardt's solution (0.1% Ficoll/0.1% polyvinylpyrrolidone/0.1% bovine serum albumin)/50 mM-sodium phosphate (pH 6.5)/yeast

RNA (200 μg/ml)/0.2% SDS/sheared and denatured E. coli DNA (100 μg/ml). Hybridization was at 42 °C for 24 h in the same buffer, except that E. coli DNA was at 300 μg/ml, poly(dC) was added to 20 μg/ml, and the cDNA probe was added to 5×10^5 c.p.m./ml. Filters were rinsed in 2 × SSC/16 mM-sodium phosphate/0.05% sodium pyrophosphate/0.1% SDS for 10 min at room temperature, and then washed essentially in the same solution, but at 65 °C and with 0.1 × SSC in the final step.

Sequencing of the selected clones

PstI or PstI/BamHI restriction fragments of the selected C1r clones were subcloned into M13 bacteriophage vectors (Messing, 1983) and sequenced by the dideoxy-chain-termination method of Sanger et al. (1977), using [α-^{35}S]dATP (Amersham International). Occasional ambiguities were resolved by sequencing appropriate fragments according to the Maxam & Gilbert's protocol (Maniatis et al., 1982).

Fig. 2. Complete sequence of the human C1r precursor deduced from cDNA sequence analyses

Nucleotide numbering starts with the first residue of the pHC1r/253 insert. Relevant stop codons are boxed. An ATTAAA polyadenylation signal is underlined, and it is followed by an arrow that marks the location of the poly(A) stretch found in clone pHC1r/5. Amino acid numbering is based on the N-terminal residue (+1) of the plasma form. The peptide that provided the information for the design of the synthetic probe is boxed. A vertical bar marks the site where the zymogen form of C1r is cleaved. The encircled residues 135 and 150 mark a site of protein heterogeneity and a site of complex post-translational modification respectively. N-Linked glycosylation sites are indicated by ★. Cysteine residues belonging to conserved sequence elements discussed in the text are represented by bold lettering. An arginine residue that may play a role in the maturation of the C1r precursor is also shown in bold lettering in the leader peptide.

```
                                                                    -17
                                                                     M   W   L   L   Y   L   L   V   P   A   L   F   C
   1  TGCACGAAGACGCTGTCGGGAGAGCCCAGGATTCAACACGGGCCT |TGA| GAA ATG TGG CTC TTG TAC CTC CTG GTG CCG GCC CTG TTC TGC
                                    +1
  -4      R   A   G   G   S   I   P   I   P   Q   K   L   F   G   E   V   T   S   P   L   F   P   K   P   Y   P   N
  91  AGG GCA GGA GGC TCC ATT CCC ATC CCT CAG AAG TTA TTT GGG GAG GTG ACT TCC CCT CTG TTC CCC AAG CCT TAC CCC AAC

  24      N   F   E   T   T   T   V   I   T   V   P   T   G   Y   R   V   K   L   V   F   Q   Q   F   D   L   E   P
 172  AAC TTT GAA ACA ACC ACT GTG ATC ACA GTC CCC ACG GGA TAC AGG GTG AAG CTC GTC TTC CAG CAG TTT GAC CTG GAG CCT

  51      S   E   G   C   F   Y   D   Y   V   K   I   S   A   D   K   K   S   L   G   R   F   C   G   Q   L   G   S
 253  TCT GAA GGC TGC TTC TAT GAT TAT GTC AAG ATC TCT GCT GAT AAG AAA AGC CTG GGG AGG TTC TGT GGG CAA CTG GGT TCT

  78      P   L   G   N   P   P   G   K   K   E   F   M   S   Q   G   N   K   M   L   L   T   F   H   T   D   F   S
 334  CCA CTG GGC AAC CCC CCG GGA AAG AAG GAA TTT ATG TCC CAA GGG AAC AAG ATG CTG CTG ACC TTC CAC ACA GAC TTC TCC

 105      N   E   E   N   G   T   I   M   F   Y   K   G   F   L   A   Y   Y   Q   A   V   D   L   D   E   C   A   S
 415  AAC GAG GAG AAT GGG ACC ATC ATG TTC TAC AAG GGC TTC CTG GCC TAC TAC CAA GCT GTG GAC CTT GAT GAA TGT GCT TCC

 132      R   S   K   L   G   E   E   D   P   Q   P   Q   C   Q   H   L   C   H   N   Y   V   G   G   Y   F   C   S
 496  CGG AGC AAA TTA GGG GAG GAG GAT CCC CAG CCC CAG TGC CAG CAC CTG TGT CAC AAC TAC GTT GGA GGC TAC TTC TGT TCC

 159      C   R   P   G   Y   E   L   Q   E   D   R   H   S   C   Q   A   E   C   S   S   E   L   Y   T   E   A   S
 577  TGC CGT CCA GGC TAT GAG CTT CAG GAA GAC AGG CAT TCC TGC CAG GCT GAG TGC AGC AGC GAG CTG TAC ACG GAG GCA TCA

 186      G   Y   I   S   S   L   E   Y   P   R   S   Y   P   P   D   L   R   C   N   Y   S   I   R   V   E   R   G
 658  GGC TAC ATC TCC AGC CTG GAG TAC CCT CGG TCC TAC CCC CCT GAC CTG CGC TGC AAC TAC AGC ATC CGG GTG GAG CGG GGC

 213      L   T   L   H   L   K   F   L   E   P   F   D   I   D   D   H   Q   Q   V   H   C   P   Y   D   Q   L   Q
 739  CTC ACC CTG CAC CTC AAG TTC CTG GAG CCT TTT GAT ATT GAT GAC CAC CAG CAA GTA CAC TGC CCC TAT GAC CAG CTA CAG

 240      I   Y   A   N   G   K   N   I   G   E   F   C   G   K   Q   R   P   P   D   L   D   T   S   S   N   A   V
 820  ATC TAT GCC AAC GGG AAG AAC ATT GGC GAG TTC TGT GGG AAG CAA AGG CCC CCC GAC CTC GAC ACC AGC AGC AAT GCT GTG

 267      D   L   L   F   F   T   D   E   S   G   D   S   R   G   W   K   L   R   Y   T   T   E   I   I   K   C   P
 901  GAT CTG CTG TTC TTC ACA GAT GAG TCG GGG GAC AGC CGG GGC TGG AAG CTG CGC TAC ACC ACC GAG ATC ATC AAG TGC CCC

 294      Q   P   K   T   L   D   E   F   T   I   I   Q   N   L   Q   P   Q   Y   Q   F   R   D   Y   F   I   A   T
 982  CAG CCC AAG ACC CTA GAC GAG TTC ACC ATC ATC CAG AAC CTG CAG CCT CAG TAC CAG TTC CGT GAC TAC TTC ATT GCT ACC

 321      C   K   Q   G   Y   Q   L   I   E   G   N   Q   V   L   H   S   F   T   A   V   C   Q   D   D   G   T   W
1063  TGC AAG CAA GGC TAC CAG CTC ATA GAG GGG AAC CAG GTG CTG CAT TCC TTC ACA GCT GTC TGC CAG GAT GAT GGC ACG TGG

 348      H   R   A   M   P   R   C   K   I   K   D   C   G   Q   P   R   N   L   P   N   G   D   F   R   Y   T   T
1144  CAT CGT GCC ATG CCC AGA TGC AAG ATC AAG GAC TGT GGG CAG CCC CGA AAC CTG CCT AAT GGT GAC TTC CGT TAC ACC ACC

 375      T   M   G   V   N   T   Y   K   A   R   I   Q   Y   Y   C   H   E   P   Y   Y   K   M   Q   T   R   A   G
1225  ACA ATG GGA GTG AAC ACC TAC AAG GCC CGT ATC CAG TAC TAC TGC CAT GAG CCA TAT TAC AAG ATG CAG ACC AGA GCT GGC

 402      S   R   E   S   E   Q   G   V   Y   T   C   T   A   Q   G   I  |W   K   N   E   Q   K   G|  E   K   I   P
1306  AGC AGG GAG TCT GAG CAA GGG GTG TAC ACC TGC ACA GCA CAG GGC ATT TGG AAG AAT GAA CAG AAG GGA GAG AAG ATT CCT

 429      R   C   L   P   V   C   G   K   P   V   N   P   V   E   Q   R   Q   R | I   I   G   G   Q   K   A   K   M
1387  CGG TGC TTG CCA GTG TGT GGG AAG CCC GTG AAC CCC GTG GAA CAG ACG CAG CGC ATC ATC GGA GGG CAA AAA GCC AAG ATG

 456      G   N   F   P   W   Q   V   F   T   N   I   H   G   R   G   G   A   L   L   G   D   R   W   I   L   T
1468  GGC AAC TTC CCC TGG CAG GTG TTC ACC AAC ATC CAC GGG CGC GGG GGC GGG GCC CTG CTG GGC GAC CGC TGG ATC CTC ACA

 483      A   A   H   T   L   Y   P   K   E   H   E   A   Q   S   N   A   S   L   D   V   F   L   G   H   T   N   V
1549  GCT GCC CAC ACC CTG TAT CCC AAG GAA CAC GAA GCG CAA AGC AAC GCC TCT TTG GAT GTG TTC CTG GGC CAC ACA AAT GTG

 510      E   E   L   M   K   L   G   N   H   P   I   R   R   V   S   V   H   P   D   Y   R   Q   D   E   S   Y   N
1630  GAA GAG CTC ATG AAG CTA GGA AAT CAC CCC ATC CGC AGG GTC AGC GTC CAC CCG GAC TAC CGT CAG GAT GAG TCC TAC AAT

 537      F   E   G   D   I   A   L   L   E   L   E   N   S   V   T   L   G   P   N   L   L   P   I   C   L   P   D
1711  TTT GAG GGG GAC ATC GCC CTG CTG GAG CTG GAA AAT AGT GTC ACC CTG GGT CCC AAC CTC CTC CCC ATC TGC CTC CCT GAC

 564      N   D   T   F   Y   D   L   G   L   M   G   Y   V   S   G   F   G   V   M   E   E   K   I   A   H   D   L
1792  AAC GAT ACC TTC TAC GAC CTG GGC TTG ATG GGC TAT GTC AGT GGC TTC GGG GTC ATG GAG GAG AAG ATT GCT CAT GAC CTC

 591      R   F   V   R   L   P   V   A   N   P   Q   A   C   E   N   W   L   R   G   K   N   R   M   D   V   F   S
1873  AGG TTT GTC CGT CTG CCC GTA GCT AAT CCA CAG GCC TGT GAG AAC TGG CTC CGG GGA AAG AAT AGG ATG GAT GTG TTC TCT

 618      Q   N   M   F   C   A   G   H   P   S   L   K   Q   D   A   C   Q   G   D   S   G   G   V   F   A   V   R
1954  CAA AAC ATG TTC TGT GCT GGA CAC CCA TCT CTA AAG CAG GAC GCC TGC CAG GGG GAT AGT GGG GGC GTT TTT GCA GTA AGG

 645      D   P   N   T   D   R   W   V   A   T   G   I   V   S   W   G   I   G   C   S   R   G   Y   G   F   Y   T
2035  GAC CCG AAC ACT GAT CGC TGG GTG GCC ACG GGC ATC GTG TCC TGG GGC ATC GGG TGC AGC AGG GGC TAT GGC TTC TAC ACC

 672      K   V   L   N   Y   V   D   W   I   K   K   E   M   E   E   E   D
2116  AAA GTG CTC AAC TAC GTG GAC TGG ATC AAG AAA GAG ATG GAG GAG GAG GAC |TGA| GCCCAGAATTCACTAGGTTCGAATCCAGAGAGC

2203  AGTGTGGAAAAAAAAAAACAAAAAACAACTGACCAGTTGTTGATAACCACTAAGAGTCTCTATTAAAATTACTGATGCAGAAAGACCGTGTGTGAAATTCT

2304  CTTTCCTGTAGTCCCATTGATGTACTTTACCTGAAACAACCAAAGGGCCCCTTTCTTTCTTCTGAGGATTGCAGAGGATATAG
```

RESULTS

cDNA clones encoding human C1r

Screening of about 40000 colonies from the initial library was performed with a mixture of 16 synthetic icosamers covering the amino acid sequence WKNEQKG (one-letter notation) (Fig. 1), a region located near the C-terminal extremity of the C1r A chain (Gagnon & Arlaud, 1985). The choice of this peptide was based on considerations of minimal codon degeneracy. Thus the synthetic oligonucleotide mixture:

$$5'd(C-C-\textcircled{C}-T-T-C/T-T-G-C/T-T-C-A/G-T-T-C/$$
$$T-T-T-C-C-A)3'$$

was designed by taking into account all codon possibilities for four of the five internal residues with codon degeneracy of two (K, N, E, Q, as underlined in the peptide sequence shown above). The encircled nucleotide, which corresponds to the wobble position of the codon for the second K, was chosen as dC on the basis of thermal-stability considerations and only the invariant positions of codons for glycine were considered.

Among the 30 positive clones isolated in the first round of screening, pHC1r/5 contained the longest insert (1.5 kb). Partial sequence analysis verified its identity and provided the alignment to the C1r protein as shown in Fig. 1.

In order to obtain full-length clones that would allow us to determine the sequence of the C1r precursor, we screened a sublibrary, which had been enriched for large cDNA inserts. To this end we used as a probe the 5'-terminal 200-bp-long endonuclease-PstI fragment of the pHC1r/5 insert, as described in the Materials and methods section. Out of 14 clones obtained, two were of particular interest: pHC1r/213 contains an insert of 1947 bp, which covers additional 3' non-coding sequences; pHC1r/253 contains an insert of 2170 bp and encodes the entire C1r precursor.

Features of the C1r cDNA sequence

This cDNA sequence was essentially obtained from pHC1r/253, except for the 3' non-coding region. However, as pHC1r/253 presents a 5-bp deletion, from position 489 to position 493, this gap was covered by sequencing a fragment of clone pHC1r/213. This deletion is most likely a cloning artefact of cDNA synthesis. The sequence shown in Fig. 2 covers a total length of 2386 nucleotides. A stretch of 51 5' non-coding nucleotides is followed by a region encoding a 17-residues-long leader peptide. The ATG codon at position 52-54 in the nucleotide sequence, corresponding to position − 17 in the protein sequence, must be the initial methionine codon. In fact, there is a TGA stop codon at position 46-48, and no other ATG codon before the serine codon encoding the N-terminal residue of the mature C1r (Arlaud et al., 1986b).

The following 2064 nucleotides completely confirm the amino acid sequence and its four glycosylation sites, as previously determined (Arlaud & Gagnon, 1983; Arlaud et al., 1986b). Surprisingly, the putative β-hydroxyaspartic acid residue located at position 150 (Arlaud et al., 1986b) is encoded as asparagine.

A comparison of the inserts of pHC1r/5 and pHC1r/213 provides interesting information concerning the 3'-end of the C1r messenger. The 116 nucleotides of

3' non-coding sequences found in pHC1r/5 include the TGA stop codon and an ATTAAA polyadenylation signal, which is followed by a poly(A) tail starting 13 nucleotides downstream. Conversely, pHC1r/213 presents an overlapping, but longer, 3' non-coding sequence (220 nucleotides long), without poly(A) tail or additional polyadenylation signals. This finding correlates with our detection of two populations of C1r messengers, one of about 2.9 kb and the second some 200 nucleotides shorter, in Northern-blot analyses of human liver or Hep G2 mRNA (results not shown).

DISCUSSION

These results confirm the peptide sequence studies on C1r (Arlaud & Gagnon, 1983; Gagnon & Arlaud, 1985) and the N-terminal sequence of its plasma form, which has been recently determined by Arlaud et al. (1986b). Our data and their finding of a serine residue at the N-terminal position allow the definition of a 17-amino-acids-long leader peptide in the C1r precursor. The presence in this leader sequence of an arginine residue at position − 4 opens the possibility that proteolytic cleavage at this site produces an intermediate precursor form of C1r retaining the additional N-terminal AGG peptide. Although such maturation intermediate lacks direct experimental evidence, it is tempting to speculate that this could explain the anomalous functional properties of the C1r protein secreted by Hep G2 cells. Reboul et al. (1986) have indeed shown that this protein fails to autoactivate, although it is susceptible to activation by exogenous C1r and displays the main features of purified plasma C1r. However, one has to consider that the maturation of C1r can be affected at a number of other relevant steps, e.g. glycosylation and/or modifications leading to the putative β-hydroxyaspartic acid residue (see below). Expression studies with cDNA or genomic clones should discriminate between these possibilities.

Heterogeneity in the polypeptide sequence of the mature C1r has been pointed out by Arlaud et al. (1986b) at position 135, where they found serine and leucine at a ratio of 2.5:1. We found that a leucine residue is encoded at this position by the two cDNA clones covering this region (pHC1r/213 and pHC1r/253). It is noteworthy that this position is part of an epidermal-growth-factor-like sequence (Doolittle, 1985) located in a region that most likely is crucial in determining the functional properties of the interaction domain of C1r. Within this epidermal-growth-factor-like sequence, a β-hydroxyaspartic acid residue at position 150 has been tentatively identified by Arlaud et al. (1986b). Our cDNA sequencing data reveal that this residue is in fact encoded as asparagine. Since aspartic acid is found encoded at the corresponding position of other β-hydroxyaspartic acid-containing proteins [bovine (Long et al., 1984) and human (Foster et al., 1985) protein C; human factor IX (Anson et al., 1984)], it appears that a more complex post-translational modification should occur at this location in C1r, involving both a deamination and a hydroxylation step.

Internal-homology regions at positions 10-78 and 186-257 of the A-chain of C1r have recently been discussed (Arlaud et al., 1986b); up to now, they have no known counterpart in other proteins, with the exception

of C1s (Spycher *et al.*, 1986; M. Tosi & C. Duponchel, unpublished work).

Recently, conserved cysteine-containing stretches of about 60 amino acids have been found in several complement proteins that share the property of interacting with C3b or C4b, but also in non-complement proteins (reviewed by Reid *et al.*, 1986). Interestingly, two such homology regions (residues 292–354 and 359–430) are located in tandem near the C-terminus of the C1r A-chain. The characteristically located four cysteine residues of each of these repeats are indicated with **bold** lettering in Fig. 2. It is also noteworthy that two such repeats are similarly located in C1s, the other enzymic subcomponent of C1 (M. Tosi & C. Duponchel, unpublished work). They might be involved in the control of C1 activity by C3b and C4b, which has been recently proposed by Ziccardi (1986), or in the recognition of C4, one of the two substrates of C1s. These two domains might also play a role in the C1r structure: they account for most of the γ-segment of the A-chain (amino acids 280–446), which has been shown to be intimately associated with the B-chain within the catalytic domains of the C1r$_2$ dimer (Arlaud *et al.*, 1986a).

REFERENCES

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A. & Brownlee, G. G. (1984) EMBO J. 3, 1053–1060

Arlaud, G. J. & Gagnon, J. (1983) Biochemistry 22, 1758–1764

Arlaud, G. J. & Gagnon, J. (1985a) FEBS Lett. 117, 289–294

Arlaud, G. J. & Gagnon, J. (1985b) Complement 2, 4

Arlaud, G. J., Villiers, C. L., Chesne, S. & Colomb, M. G. (1980) Biochim. Biophys. Acta 616, 116–129

Arlaud, G. J., Gagnon, J., Villiers, C. L. & Colomb, M. G. (1986a) Biochemistry 25, 5177–5182

Arlaud, G. J., Willis, A. C. & Gagnon, J. (1986b) Biochem. J., in the press

Cooper, N. R. (1985) Adv. Immunol. 37, 151–216

Doolittle, R. F. (1985) Trends Biochem. Sci. 10, 233–237

Foster, D. C., Yoshitake, S. & Davie, E. W. (1985) Proc. Natl. Acad. Sci. U.S.A. 82, 4673–4677

Gagnon, J. & Arlaud, G. J. (1985) Biochem. J. 225, 135–142

Gubler, U. & Hoffman, B. J. (1983) Gene 25, 263–269

Hanahan, D. (1983) J. Mol. Biol. 166, 557–580

Hubacek, J. & Glover, S. W. (1970) J. Mol. Biol. 50, 111–127

Long, G. L., Belagaje, R. M. & McGillivray, R. T. A. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 5653–5656

Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Messing, J. (1983) Methods Enzymol. 101, 20–78

Reboul, A., Bensa, J.-C. & Colomb, M. G. (1986) Biochem. J. 233, 559–564

Reid, K. B. M. & Porter, R. R. (1981) Annu. Rev. Biochem. 50, 433–464

Reid, K. B. M., Bentley, D. R., Campbell, R. D., Chung, L. P., Sim, R. B., Kristensen, T. & Tack, B. F. (1986) Immunol. Today 7, 230–234

Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74 5463–5467

Spycher, S. E., Nick, H. & Rickli, E. E. (1986) Eur. J. Biochem. 156, 49–57

Tosi, M., Journet, A., Colomb, M. G. & Meo, T. (1985) Complement 2, 79

Tosi, M., Duponchel, C., Bourgarel, P., Colomb, M. G. & Meo, T. (1986) Gene 42, 265–272

Vieira, J. & Messing, J. (1982) Gene 19, 259–268

World Health Organisation (1968) Bull. W.H.O. 39, 935–936

Ziccardi, R. J. (1986) J. Immunol. 136, 3378–3385